

Appl Microbiol Biotechnol (2010) 87:479–489
DOI 10.1007/s00253-010-2628-x

MINI-REVIEW

Inteins, valuable genetic elements in molecular biology and biotechnology

Skander Elleuche · Stefanie Pöggeler

Received: 11 March 2010 / Revised: 14 April 2010 / Accepted: 14 April 2010 / Published online: 7 May 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Inteins are internal protein elements that self-excise from their host protein and catalyze ligation of the flanking sequences (exteins) with a peptide bond. They are found in organisms in all three domains of life, and in viral proteins. Intein excision is a posttranslational process that does not require auxiliary enzymes or cofactors. This self-excision process is called protein splicing, by analogy to the splicing of RNA introns from pre-mRNA. Protein splicing involves only four intramolecular reactions, and a small number of key catalytic residues in the intein and exteins. Protein-splicing can also occur in *trans*. In this case, the intein is separated into N- and C-terminal domains, which are synthesized as separate components, each joined to an extein. The intein domains reassemble and link the joined exteins into a single functional protein. Understanding the *cis*- and *trans*-protein splicing mechanisms led to the development of intein-mediated protein-engineering applications, such as protein purification, ligation, cyclization, and selenoprotein production. This review summarizes the catalytic activities and structures of inteins, and focuses on the advantages of some recent intein applications in molecular biology and biotechnology.

Keywords Intein · Protein splicing · Protein purification · Cyclic proteins · Selenoproteins

Introduction

Inteins were identified 20 years ago, when two groups reported an in-frame insertion in the *VMA1* gene, which encodes a vacuolar membrane H⁺-ATPase of the yeast *Saccharomyces cerevisiae* (Hirata et al. 1990; Kane et al. 1990). The nucleotide sequence of the *VMA1* gene predicts a polypeptide of 1,071 amino acids with a calculated molecular mass of 118 kDa, but the size of the VMA1 protein, as estimated from sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gels, is only 67 kDa. Furthermore, the N- and C-terminal regions of the deduced sequence were shown to be very similar to the catalytic subunits of vacuolar membrane H⁺-ATPases of other organisms, while an internal region of 454 amino acid residues displayed no detectable sequence similarity to any known ATPase subunits. Instead, the internal sequence exhibits similarity to an *S. cerevisiae* endonuclease encoded by the *HO* gene. The in-frame insertion was found to be present in the mRNA, translated with the Vma1 protein, and excised posttranslationally (Kane et al. 1990). By analogy to pre-mRNA introns and exons, the segments are called *intein* for internal protein sequence, and *extein* for external protein sequence, with upstream exteins termed N-exteins and downstream exteins called C-exteins. The post-translational process that excises the internal region from the precursor protein, with subsequent ligation of the N- and C-exteins, is termed protein splicing (Perler et al. 1994). The products of the protein splicing process are two stable proteins, the mature protein and the intein (Fig. 1). According to accepted nomenclature, intein names include a genus and species designation, abbreviated with three letters, and a host

S. Elleuche
Institute of Technical Microbiology, Technical University
Hamburg-Harburg,
Kasernenstr. 12,
21073 Hamburg, Germany

S. Pöggeler (✉)
Institute of Microbiology and Genetics,
Department Genetics of Eukaryotic Microorganisms,
Georg-August-University of Göttingen,
Grisebachstr. 8,
37077 Göttingen, Germany
e-mail: spoegge@gwdg.de

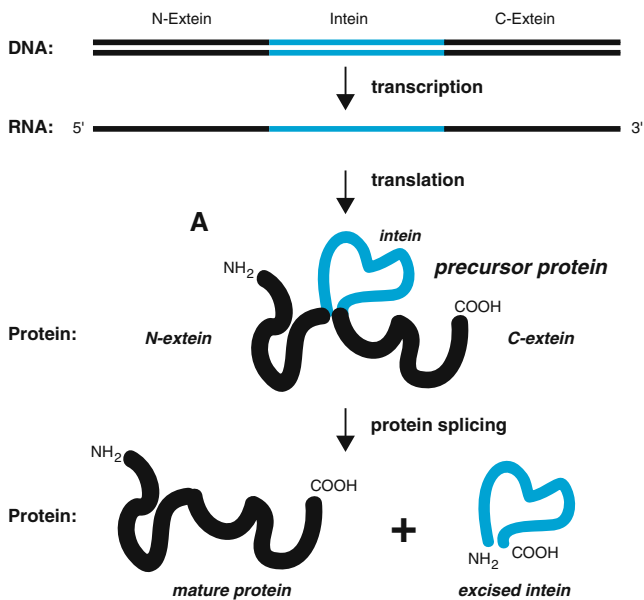


Fig. 1 Protein splicing. The intein coding sequence is transcribed into mRNA and translated to a nonfunctional protein precursor, which then undergoes a self-catalyzed rearrangement in which the intein is excised and the exteins are joined to yield the mature protein

gene designation. For example, the *S. cerevisiae* VMA1 intein is called *Sce* VMA1. Multiple inteins from one protein are numbered with Arabic numerals (Perler 2002). Large-scale genome sequencing approaches have identified inteins in all three domains of life, as well as in phages and viruses. By the end of 2009, the intein registry InBase at <http://www.neb.com/neb/inteins.html> (Perler 2002) listed more than 450 inteins in the genomes of Eubacteria, Archaea, and Eukarya. In prokaryotes, intein sequences often reside within proteins involved in DNA replication, repair, or transcription, such as DNA and RNA polymerases, RecA, helicases, or gyases, and in the cell division control protein CDC21. Others are located in metabolic enzymes including ribonucleoside triphosphate reductase, and UDP-glucose dehydrogenase (Perler 2002; Starokadomskyy 2007). Eukaryotic inteins are encoded in the nuclear genes of fungi, and in the nuclear or plastid genes of some unicellular algae. In fungi, intein sequences are found in homologs of the *S. cerevisiae* VMA1 gene or in the *prp8* genes, but they are also found in genes encoding glutamate synthases, chitin synthases, threonyl-tRNA synthetases, and subunits of DNA-directed RNA polymerases (Elleuche and Pöggeler 2009; Poulter et al. 2007). In green and cryptophyte algae, inteins reside within the chloroplast ClpP protease, the RNA polymerase beta subunit, the DnaB helicase and the nuclear RNA polymerase II (Douglas and Penny 1999; Luo and Hall 2007; Turmel et al. 2008; Wang and Liu 1997).

Most genes encode only one intein, and inteins found at the same insertion site in homologous extein genes are considered intein alleles (Perler et al. 1997). In rare cases,

genes encode more than one intein, such as the ribonucleotide reductase gene of the oceanic N₂-fixing cyanobacterium *Trichodesmium erythraeum*, which encodes four inteins (Liu et al. 2003).

Structure of mini-inteins and large inteins

Inteins are classified into two groups, large and minimal (mini) (Liu 2000). Large inteins contain a homing endonuclease domain that is absent in mini-inteins. Homing endonucleases are site-specific, double-strand DNA endonucleases that promote the lateral transfer between genomes of their own coding region with flanking sequences, in a recombination-dependent process known as "homing." Usually, homing endonucleases are encoded by an open reading frame within an intron or intein (Belfort et al. 2005; Chevalier and Stoddard 2001). Large inteins are bi-functional proteins, with a protein splicing domain, and a central endonuclease domain. Splicing-efficient mini-inteins have been engineered from large inteins by deleting the central endonuclease domain, demonstrating that the endonuclease domain is not involved in protein splicing (Chong and Xu 1997; Derbyshire et al. 1997; Shingledecker et al. 1998). The splicing domain is split by the endonuclease domain into N- and C-terminal subdomains, which contain conserved blocks of amino acids, with blocks A, N2, B, and N4 in the N-terminal subdomain, and blocks G and F in the C-terminal subdomain (Perler et al. 1997; Pietrokovski 1994, 1998). These domains can also be identified in mini-inteins (Fig. 2). The three-dimensional structures of naturally occurring mini-inteins and engineered mini-inteins reveal that the N- and C-terminal splicing domains form a common horseshoe-like 12-β-strand scaffold termed the Hedgehog/Intein (HINT) module (Ding et al. 2003; Hall et al. 1997; Klabunde et al. 1998; Koonin 1995; Perler 1998; Sun et al. 2005; Van Roey et al. 2007).

All known inteins share a low degree of sequence similarity, with conserved residues only at the N- and C-termini. Most inteins begin with Ser or Cys and end in His-Asn, or in His-Gln. The first amino acid of the C-extein is an invariant Ser, Thr, or Cys, but the residue preceding the intein at the N-extein is not conserved (Perler 2002). However, residues proximal to the intein-splicing junction at both the N- and C-terminal exteins were recently found to accelerate or attenuate protein splicing (Amitai et al. 2009).

Cis- and trans-splicing mechanisms of inteins

Protein splicing is a rapid process of four nucleophilic attacks, mediated by three of the four conserved splice junction residues. In step 1, the splicing process begins with

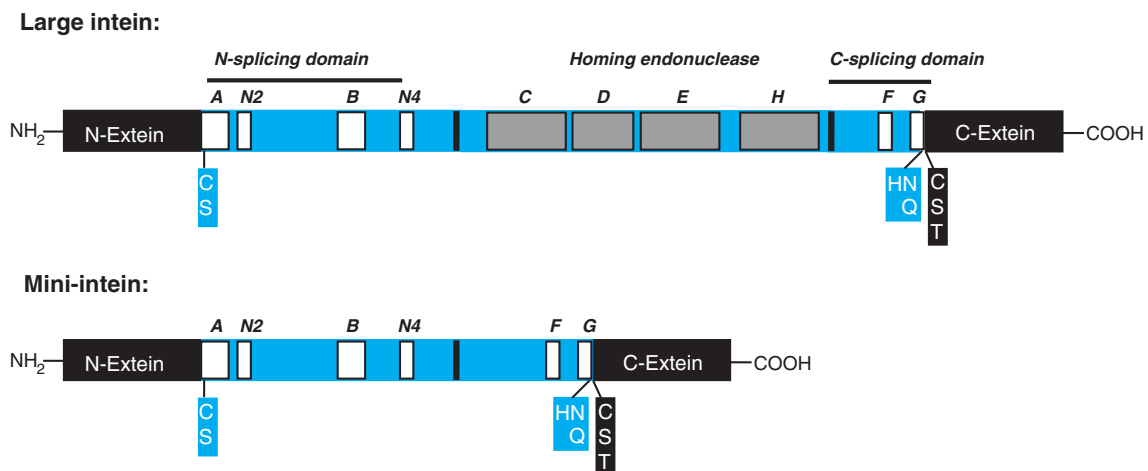


Fig. 2 Structure of large and mini-inteins. Conserved elements in a large intein and mini-intein are indicated. The white and grey areas A, N2, B, N4, C, D, E, H, F, and G are conserved intein motifs identified by Pietrovski (1994, 1998) and Perler et al. (1994). The exteins are

illustrated in black and the intein sequence in blue. The site of insertion of the homing endonuclease in large inteins is indicated by the dark vertical line. Conserved amino acid residues of the intein and the C-extein are indicated below

an N–O shift if the first intein residue is Ser, or N–S acyl shift, if the first intein residue is Cys. This forms a (thio) ester bond at the N-extein/intein junction. In step 2, the (thio)ester bond is attacked by the OH- or SH-group of the first residue in the C-extein (Cys, Ser, or Thr). This leads to a transesterification, which transfers the N-extein to the side-chain of the first residue of the C-extein. In step 3, the cyclization of the conserved Asn residue at the C-terminus of the intein releases the intein and links the exteins by a (thio)ester bond. Finally, step 4 is a rearrangement of the (thio)ester bond to a peptide bond by a spontaneous S–N or O–N acyl shift (Fig. 3). Details of the chemical process involved in protein splicing of standard (class I inteins) have been comprehensively described and reviewed (Gogarten et al. 2002; Liu 2000; Noren et al. 2000; Paulus 2000; Saleh and Perler 2006; Starokadomskyy 2007; Tori et al. 2010). In addition to the standard protein splicing pathway, new classes of inteins performing an alternative splicing process have been described recently. These inteins lack the N-terminal Ser or Cys residue and are classified as class 2 and class 3 inteins (Southworth et al. 2000; Tori et al. 2010). Both cannot perform the acyl shift that initiates the splicing reaction in class 1 inteins. In class 2 inteins, present in archaeal KlbA proteins, the first residue of the C-extein (nucleophile Cys) directly attacks the amide bond at the N-terminal splice site junction to form a standard branched intermediate (Johnson et al. 2007; Southworth et al. 2000). In the class 3 intein of the mycobacteriophage Bethlehem DnaB protein, a Cys residue of the conserved block F attacks the peptide bond at the N-terminal splice site junction, forming a branched intermediate with a labile thioester linkage. The N-extein is then transferred by a transesterification to the first residue of the C-extein (Thr),

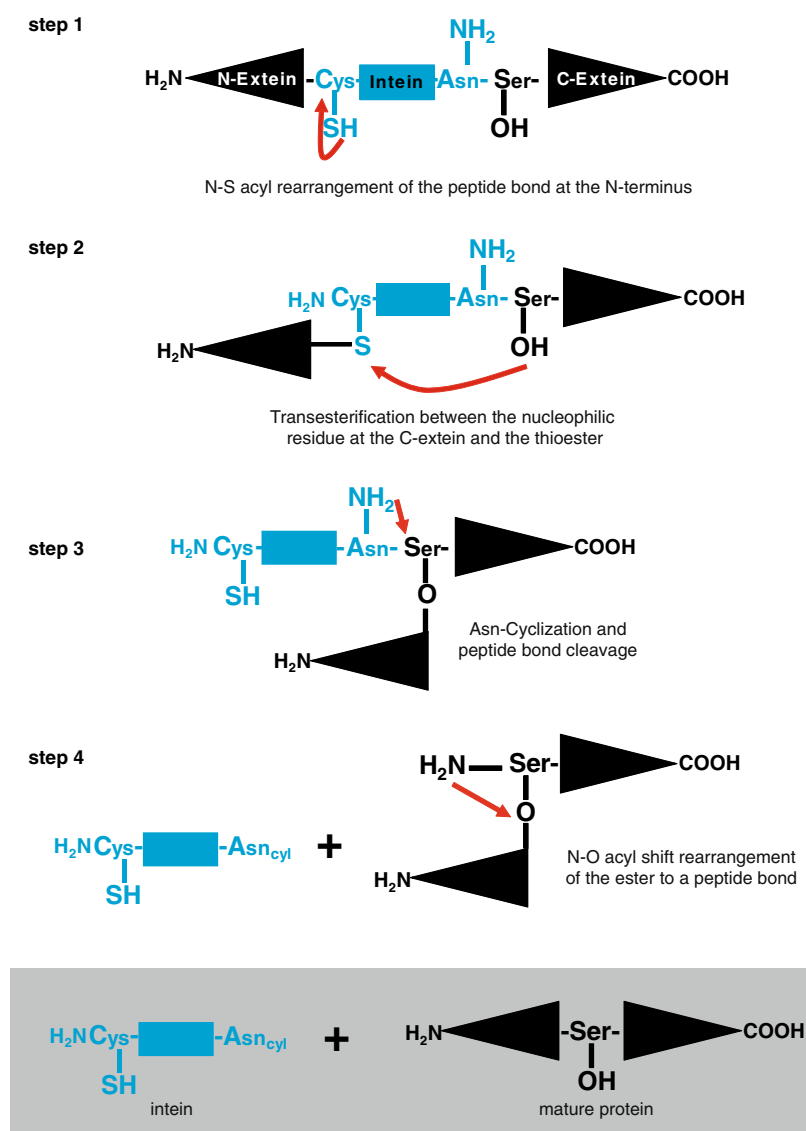
which results in the formation of a standard branched intermediate as in class 1 inteins (Tori et al. 2010).

Site-specific cleavage of the intein–extein junctions in class 1 inteins can be achieved by mutation of the conserved intein residues. Mutation of the Asn residue at the intein C-terminus abolishes steps 3 and 4 of the splicing reaction and results in N-terminal cleavage. Since step 1 still occurs, the (thio)ester bond can spontaneously hydrolyze, separating the N-extein from the intein/C-extein portion. Mutation of the conserved first residue of the intein abolishes steps 1, 2, and 4 of the splicing reaction and leads to C-terminal cleavage. In such a mutated intein, Asn cyclization (step 3) still occurs, to separate the C-extein from the N-extein/intein portion. Controllable cleavage of modified *cis*-splicing inteins has been adapted for a wide range of useful applications in molecular biology and biotechnology (see below).

Interestingly, inteins can also exist as two fragments encoded by two separately transcribed and translated genes. These so-called split inteins self-associate and catalyze protein-splicing activity in *trans*. The first native split intein capable of protein *trans*-splicing was identified in the cyanobacterium *Synechocystis* sp. strain PCC6803. The N- and C-terminal halves of the *Synechocystis* catalytic subunit alpha of DNA polymerase III DnaE are encoded by the *dnaE-n* and *dnaE-c* genes, which are more than 700 kb apart (Wu et al. 1998).

Split inteins have been identified in diverse cyanobacteria and archaea (Caspi et al. 2003; Choi et al. 2006; Dassa et al. 2007; Liu and Yang 2003; Wu et al. 1998; Zettler et al. 2009), but have not been found in eukaryotes thus far. Recently, a bioinformatic analysis of environmental metagenomic data revealed 26 different loci with a novel genomic

Fig. 3 Splicing mechanism of inteins. Inteins take place in four reaction steps. See text for details



arrangement. At each locus, a conserved enzyme coding region is interrupted by a split intein, with a free-standing endonuclease gene inserted between the sections coding for intein subdomains. This fractured gene organization appears to be present mainly in phages (Dassa et al. 2009).

Trans-splicing of inteins can also be artificially engineered from *cis*-splicing bacterial and fungal inteins (Elleuche and Pöggeler 2007; Mills et al. 1998; Mootz and Muir 2002; Southworth et al. 1998). This is achieved mainly by separating naturally or artificially split inteins between motifs B and F, resulting in N-terminal intein fragments (I_N) of 70–110 amino acids, and C-terminal intein fragments (I_C) of ~40 amino acids. Inteins can also be artificially split at other sites. The *Ssp* DnaB mini-intein can be split at different loop regions between β -strands, yet still maintain the ability to splice in *trans*. Even an intein split into three pieces can function in protein *trans*-splicing, even if one piece is only an 11-amino acid I_N (Sun et al.

2004). Naturally split inteins and engineered split inteins can be used in various applications. In the following sections, we will briefly summarize some of the many uses of inteins as molecular biology and biotechnology tools.

Applications of inteins in biotechnology

Inteins are valuable tools in a wide range of biotechnological applications. The ligation of peptides and proteins using the natural splicing activity of inteins is known as intein-mediated protein ligation (IPL), or expressed protein ligation (EPL), and is already well established in molecular biology and biotechnology methods (Evans et al. 1999; Muir et al. 1998; Severinov and Muir 1998). Furthermore, inteins have been used for segmental labeling of proteins for NMR analysis, cyclization of proteins, controlled expression of toxic proteins, conjugation of quantum dots to proteins, and incorporation of

non-canonical amino acids (Arnold 2009; Charalambous et al. 2009; Oemig et al. 2009; Seyedsayamdost et al. 2007; Züger and Iwai 2005). In basic research studies, they have been used to monitor *in vivo* protein–protein interactions, or specifically translocate proteins into cellular organelles (Chong and Xu 2005; Ozawa and Umezawa 2005; Ozawa et al. 2003, 2005). Most of the inteins used in biotechnology are derived from prokaryotic organisms, or are engineered variants of the *S. cerevisiae* VMA1-intein.

Intein-mediated protein purification

Isolation of large amounts of highly purified proteins is a major task in biotechnology. The development of a wide range of affinity tags has greatly simplified the separation of recombinant proteins from crude extracts. Conventional affinity tags are fused as a tag sequence at the DNA level. Originally developed to isolate proteins on affinity columns or beads, and to detect proteins by Western blot, the fusion of different tag proteins and peptides can also improve the solubility and folding of the target protein (Terpe 2003; Waugh 2005). Furthermore, N-terminal tags have the advantage of enabling the efficient translation of a recombinant protein, by providing a reliable ribosome initiation site. A drawback of this approach is that the affinity tag often needs to be cleaved off the fusion protein by proteolysis with a site-specific endoprotease. For industrial applications, the removal of the affinity tag by endoproteases is the most costly step in protein production, and can interfere with the biological activity of the purified component (Wood et al. 2005). Therefore, intein-mediated bioseparation has become an excellent vehicle for affinity-tag-based protein purification techniques, and is an alternative to conventional cleavage by site-specific endoproteases.

The potential of intein-facilitated purification for a variety of proteins has been described in dozens of reports (Bastings et al. 2008; Chong et al. 1997; Gillies et al. 2008; Liu et al. 2008; Sharma et al. 2006; Singleton et al. 2002; Srinivasa Babu et al. 2009; Zhao et al. 2008).

Intein-mediated protein purification began in 1997 as a new field in biotechnology, when the *Sce* VMA1 intein was engineered to be used in purification of several prokaryotic and eukaryotic proteins (Chong et al. 1997). In principle, the exteins of an engineered intein are exchanged between the purification tag and the target protein. As described above, complete splicing of the intein can be inhibited by mutation of conserved residues at the splice junction fused to the affinity tag. This results in site-specific cleavage only at the intein–target protein border.

The immature precursor protein is usually produced in a heterologous host, and protein crude extracts are loaded on

an affinity column. After immobilization of the engineered protein and washing, the N- or C-terminal cleavage reaction is induced by either a strong nucleophile such as dithiothreitol (DTT) for the N-terminal cleavage, or a pH or temperature shift for the C-terminal cleavage (Chong et al. 1997; Mathys et al. 1999; Wood et al. 1999).

The intein-mediated purification with affinity chitin-binding tag (IMPACT) system is commercially available from New England Biolabs. This system uses a modified *Sce* VMA1 intein, fused at its C-terminus to the chitin-binding domain (CBD), and at its N-terminus to the protein of interest (Chong et al. 1997) (Fig. 4). Mutation of the C-terminal reactive Asn to Ala in the intein blocks the splicing reaction after the N–S acyl shift, and prevents C-terminal cleavage. The fusion protein accumulates as an unspliced precursor, and is purified by absorption to a chitin resin. After the addition of thiols, which serve as reactants for the transesterification reaction, N-terminal cleavage is initiated. This leads to release of the target protein with an activated thioester at the C-terminus, while the intein-CBD remains bound to the column. A new version of the IMPACT system (IMPACT CN) allows the fusion of a self-cleavable intein tag to either the C-terminus or the N-terminus of a target protein. In contrast to the first generation IMPACT-vectors, pTWIN-vectors (New England Biolabs) contain two inteins that can be used separately for protein purification, either by fusing the protein of interest to the C-terminus of one intein, or to the N-terminus of a second intein. The inteins can also be used in combination to purify a single protein by independent regulation of the cleavage reactions of intein 1 and intein 2. This method has been demonstrated using green fluorescent protein (Zhao et al. 2008).

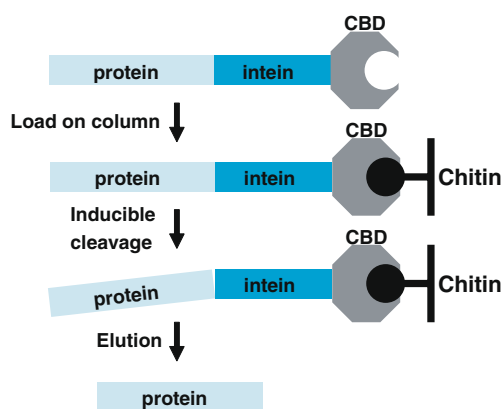


Fig. 4 Intein-mediated protein purification. Expression and purification of a protein fused to the N-terminus of a mutated intein variant. The intein is further fused to a chitin binding domain (CBD) at its C-terminus. Purification of the fusion construct is accomplished by the utilization of a chitin column. The intein is mutated at the C-terminus (Asnⁿ), so that cleavage only occurs at the N-terminus, which results in the release of the target protein from the column while the intein-CBD tag remains bound to chitin

Intein-mediated protein purification using non-chromatographic tags

An alternative to classical tagging systems that does not require expensive affinity resins is intein-mediated protein purification with new tags developed for non-chromatographic purification techniques.

One example is the combination of elastin-like polypeptides (ELP) and self-splicing inteins (Wu et al. 2006). Intein-independent protein purification methods using ELP were developed more than 10 years ago (Meyer and Chilkoti 1999). Under high-salt conditions at about 30 °C, ELP reversibly self-associates and forms insoluble aggregates that can be precipitated by centrifugation (Wu et al. 2006). A target protein fused to an intein–ELP fusion can be separated from a crude protein extract by repeated aggregation and centrifugation cycles. Finally, inducible intein cleavage enables recovery of the target protein from the intein–ELP moiety, leading to a highly pure elution fraction (Wu et al. 2006). The aggregation characteristics of ELP can be controlled by temperature, concentration, type of salt, or polypeptide length (Floss et al. 2009; Fong et al. 2009). The intein–ELP approach was recently demonstrated to be a low-cost, convenient, and potential way of generating small antimicrobial peptides (Shen et al. 2010).

Another non-chromatographic purification approach takes advantage of a multiple phasin tag, produced by bacterial species like *Cupriavidus necator* (formerly known as *Alcaligenes eutrophus* and *Ralstonia eutropha*), that specifically binds to polyhydroxybutyrate granules (PHB) in *Escherichia coli* (Banki et al. 2005; Wieczorek et al. 1995). Polyhydroxyalkanoic acids are naturally produced as a wide range of storage polymers, by several bacterial species. Their biosynthesis genes can be heterologously expressed in *E. coli*, producing polymeric granules that can be used in place of classical affinity resins for protein purification (Choi et al. 1998). The co-production in *E. coli* of phasin-tagged proteins and PHB granules enables the easy separation of the tagged target protein from crude extract, after cell lysis and centrifugation. In an impressive advancement of this system, inducing cleavage of an engineered intein releases the untagged target protein from an intein–phasin moiety, and from the bound PHB granules (Banki et al. 2005; Georgiou and Jeong 2005). In a system invented by Wood and coworkers, a pH- or temperature-inducible *Mtu* RecA intein from *Mycobacterium tuberculosis* is used for phasin tag-mediated protein purification in *E. coli*, and the authors note that the system is easily applicable to a wide range of host systems (Banki et al. 2005; Gillies et al. 2009).

The ELP and PHB systems are both highly flexible, and function efficiently with a variety of proteins, under many different conditions (Banki et al. 2005; Ge et al. 2005; Wu et al. 2006). Furthermore, both systems have been adapted

for the Gateway cloning system (Invitrogen), for rapid and easy characterization of a gene product using different vector systems (Gillies et al. 2008). The Gateway system generates a single Entry clone, from which the gene of interest is introduced directly, by simple recombination, into a number of different vectors. In addition to intein-mediated ELP and phasin fusion-protein purification, the Gateway system has been adapted for intein-mediated protein purification using classical affinity tags like maltose binding protein and CBD (Gillies et al. 2008).

Intein-mediated protein purification in large-scale processes

The use of intein-mediated procedures in bioseparation is well established at the laboratory scale and is attracting increasing interest in biotechnology. The potential of these protein purification techniques for large-scale protein production is clear, but intein-mediated protein purification systems under industrial, scaled-up conditions must be developed. The simplicity of intein-mediated protein purification, with its few purification steps and low requirement for agents, suggests that scale-up approaches have the potential to be economical in the future. Since intein-mediated cleavage does not require further downstream processing, it reduces the costs from expensive protease enzymes. Wood and co-workers designed a hypothetical scale-up method based on the DTT-inducible IMPACT system, and identified the Tris–HCl reaction buffer and the thiol compound to be the most costly ingredients in this process. They suggested exchanging the buffer system with a cheaper phosphate buffer. Cleavage induction by chemical compounds could be circumvented using inteins that are induced by physical changes (Wood et al. 2005). Furthermore, the use of non-chromatographic affinity tags could eliminate the need for expensive columns, and may also be easy to scale up. The development of recently published vectors based on Invitrogen's Gateway cloning system will facilitate production of a target protein fused to four different protein tags. Using this system, a target protein can be easily tested with different tags in a high-throughput manner (Gillies et al. 2008). In a recent review, Fong et al. (2010) describe various non-chromatographic self-cleaving purification tags and their potential industrial applications.

Self-circularization by inteins

The generation of cyclic peptides is a rapidly growing field in molecular biology and chemistry. Several methods have been established producing cyclic proteins that are exceptionally stable to chemical, thermal, or enzymatic degradation, and exhibit a higher specific activity in circular form. Increased

stability is achieved from the resistance to exoproteases, which are not capable of degrading cyclic peptides. A variety of organisms produce circular peptides with a variety of bioactivities, including anti-bacterial, uterotonic, haemolytic, and cytotoxic activity (Craik 2006). For example, the filamentous ascomycete *Tolyposcladium inflatum* produces the well-known cyclosporin A, which has long been used as an immunosuppressive drug (Thali 1995). Cyclosporin A and other known cyclic peptides are small rings of fewer than a dozen amino acids, and are produced by multidomain enzymes called peptide synthetases (Billich and Zocher 1987; Weber et al. 1994). However, many other circular proteins are synthesized as linear chains of amino acids, with the amino terminus of one residue linked to the carboxyl terminus of the next. These include cyclotides, a family of bioactive proteins from plants that contain a head-to-tail backbone and a knotted arrangement of three disulfide bonds (Craik et al. 1999).

Generally, cyclic antibiotics display an increased activity and stability in comparison to their linear analogs. Increased stability is derived from the resistance to exoproteases, which are not capable of degrading cyclic peptides. Furthermore, precisely designed small cyclic peptides can have similar specificity as endogenously produced antibiotics (Cheriyian and Perler 2009).

Inteins play a growing role in the production of cyclic peptides through the aforementioned IPL technique (Evans et al. 1999). The protein of interest is fused through its C-terminus, to the N-terminus of an intein, in which the C-terminal Asn has been mutated to be incapable of cleaving the C-terminal binding tag. The N-terminus of the target protein is altered so the second residue after the Met is a Cys. After purification of the target protein, the N-terminal Met is removed using methionyl-aminopeptidase from *E. coli*, resulting in an N-terminal Cys residue (Sancheti and Camarero 2009; Tavassoli et al. 2005). After purification, including elution from the column by thiol-induced N-terminal cleavage of the intein, the linear peptide contains a C-terminal thioester and a Cys at its N-terminus that can react to form a new peptide bond.

The pTWIN vector of the Two INtein system contains two engineered inteins (Evans and Xu 1999). The mutated *Synechocystis* sp. *Ssp* DnaB intein allows C-terminal cleavage, while *Mycobacterium xenopus* *Mxe* GyrA intein undergoes N-terminal cleavage. The combination of both proteins fused to the N- and C-terminal ends of a target protein enables the production of an N-terminal Cys residue and an activated thioester at the C-terminus, which react, resulting in cyclization (Fig. 5a). A disadvantage to this method is the low cleavage efficiency of the *Ssp* DnaB intein, which is influenced by the second and third amino acid residues following the required Cys at the N-terminus of the target protein. The introduction of a non-native linker

sequence improves cleavage efficiency, but also has the potential to interfere with the biological activity of the cyclic protein. Another problem is the possibility of polymerization instead of cyclization by activated peptides (Xu and Evans 2001).

Split inteins have also been applied in the generation of cyclic proteins and peptides. The very timely and elegant Split Intein-mediated Circular Ligation Of Peptides a ProteinS (SICLOPPS) system uses the naturally split *Synechocystis* sp. *Ssp* DnaE intein, which is fused in a rearranged order (I_C -target protein- I_N), allowing the efficient cyclization of the target protein by reconstitution of the *Ssp* DnaE intein (Fig. 5b). Using this method, it was possible to generate cyclic peptides that are short as eight amino acid residues. SICLOPPS has been used in inhibitor studies for the rapid synthesis of very large cyclic peptide libraries that are superior to the traditional chemically generated libraries, and which can be screened in vivo for new potent therapeutic drugs (Scott et al. 1999; Tavassoli and Benkovic 2007). In recent years, SICLOPPS has been impressively used, for instance, to identify several inhibitors for the dimerization of ribonucleotide reductase and 5-aminoimidazole-4-carboxamide-ribotide transformylase (for a recent review, see Cheriyian and Perler 2009).

Inteins in selenoprotein production

The 21st amino acid selenocysteine (Sec) is encoded by a UGA codon in several prokaryotic and eukaryotic proteins. Sec is incorporated during translation in a process known as recoding (Driscoll and Copeland 2003). Many selenoproteins are selenoenzymes with a single Sec residue in the active site. Since prokaryotes and eukaryotes have different UGA recoding machineries, producing selenoproteins and analyzing the characteristics of selenoenzymes in heterologous hosts is challenging (Hondal 2009). The first attempts to produce Sec-containing mammalian thioredoxin reductase (TrxR) heterologously, were undertaken in *E. coli*. In the heterologous host, UGA codes for Sec only when a specific stem-loop structure called the selenocysteine insertion sequence element (SECIS) is present in the mRNA template in close proximity to UGA, and the transacting factors SelA-D is also synthesized in the cell. Since the Sec residue of mammalian TrxR is close to the C-terminus, a SECIS element was cloned at the 3'-end of the mammalian gene (Arnér et al. 1999).

Incorporation of internal Sec residues into heterologous proteins is achieved using native chemical ligation (NCL), the related IPL, or chemical conversion of reactive Ser residues. The NCL technique facilitates the synthesis of moderately sized proteins by ligation of a peptide with a reactive thioester at the C-terminus, and a second peptide

Fig. 5 Self-circularization of peptides. **a** Self-circularization of proteins using the TWIN-system. A target protein is embedded between two intein sequences, which are modified for N- or C-terminal cleavage, respectively. The inducible splicing reaction of the inteins leads to the generation of an activated thioester residue and an N-terminal Cys for the spontaneous circularization of the linear peptide. **b** Utilization of the Split Intein-mediated Circular Ligation Of Peptides a ProteinS (SICLOPPS) also enables the circularization of peptides. In this system, the order of the naturally split *Synechocystis* sp. *Ssp* DnaE intein is inverted (I_C -target protein- I_N), and the reconstitution of the *Ssp* DnaE intein allows the efficient cyclization of the target protein

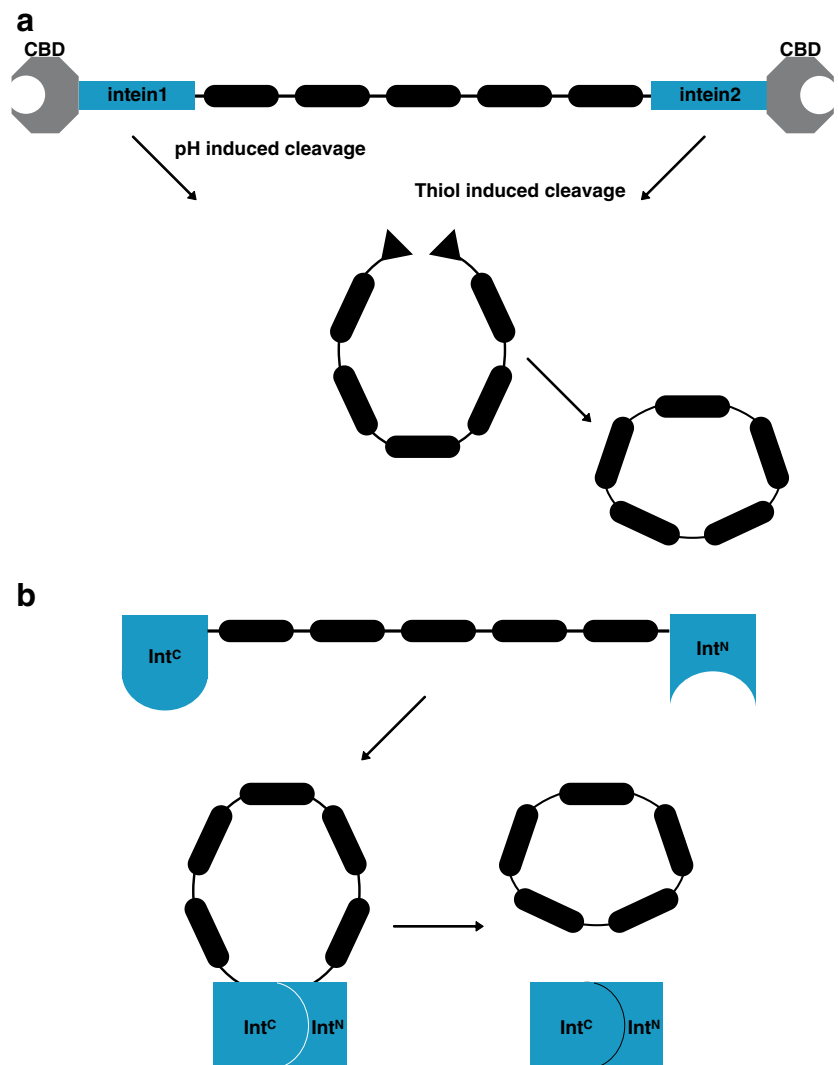
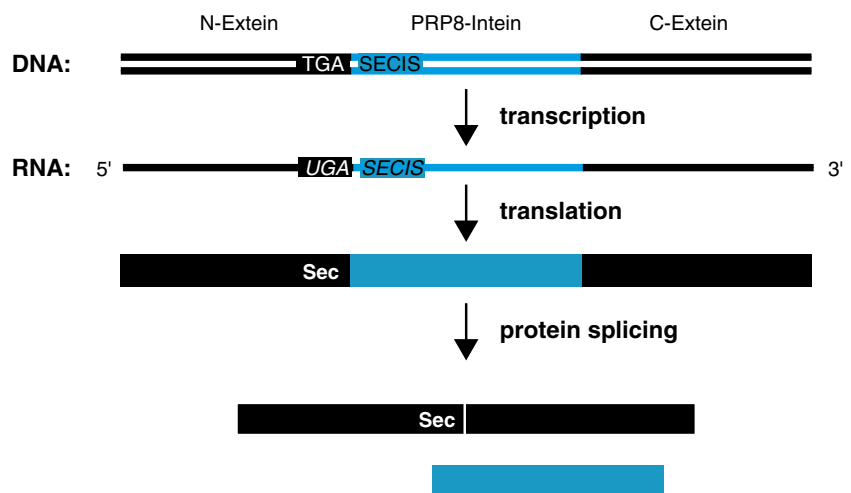


Fig. 6 Intein-mediated production of Sec-containing proteins. A UGA-codon is translated to a Sec, when a SECIS element is in close proximity at the N-terminal part of the modified fungal *Pch* PRP8-intein. The incorporation of the Sec residue followed by the protein splicing process leads to the production of a mature Sec-containing protein



containing a Cys or a Sec at the N-terminus (Dawson and Kent 2000; Hondal 2009). For IPL, the Sec-containing module is synthetically produced, while the Sec-less protein moiety is synthesized as a recombinant protein in a heterologous host, and is purified and activated by intein-mediated protein purification (Hondal 2009).

A new invention in Sec-protein production named sectein has recently been patented by the Arner group (Arnér et al. 2009). The sectein system couples expression of an intein sequence with a bacterial SECIS element and combines the advantages of SECIS elements with protein splicing, for a process that is independent of the Sec position or selenoprotein size (Fig. 6). Unlike the IPL method, chemical production of a Sec moiety is not required. The N-terminus of a selenoprotein containing the UGA Sec codon at its 3'-end acts as an N-extein. It is fused to the *Penicillium chrysogenum* Pch PRP8 intein containing a SECIS element at its 5'-end (Arnér et al. 2009; Elleuche et al. 2006). The SECIS element directs the incorporation of Sec into the peptide during translation. The C-terminus of the selenoprotein acts as the C-extein and is fused to the C-terminus of the intein. After translation, the precursor protein has a Sec residue in the N-extein, directed by SECIS element in the intein. Through protein splicing, the Sec-containing N-extein is fused to the C-extein and a mature selenoprotein is formed. The SECIS element is excised with the intein (Arnér et al. 2009).

Outlook

Since their discovery 20 years ago, the application of natural and artificial inteins has become a new and rapidly growing field in molecular biology. Protein splicing not only enriches the possibilities of posttranslational processing, but also has many prospects for applications. The protein splicing process as a protein engineering tool will become more widespread in industrial applications. The challenge is to scale up and optimize intein-mediated techniques, making them applicable and economically attractive for biotechnological processes.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M (2009) Modulation of intein activity by its neighboring extein substrates. *Proc Natl Acad Sci U S A* 106:11005–11010
- Arnér ES, Sarioglu H, Lottspeich F, Holmgren A, Bock A (1999) High-level expression in *Escherichia coli* of selenocysteine-containing rat thioredoxin reductase utilizing gene fusions with engineered bacterial-type SECIS elements and co-expression with the *selA*, *selB* and *selC* genes. *J Mol Biol* 292:1003–1016
- Arnér ES, Cheng Q, Donald HJ (2009) Method for producing selenoproteins. World Intellectual Property Organization (WIPO), United Nations
- Arnold U (2009) Incorporation of non-natural modules into proteins: structural features beyond the genetic code. *Biotechnol Lett* 31:1129–1139
- Banki MR, Gerngross TU, Wood DW (2005) Novel and economical purification of recombinant proteins: intein-mediated protein purification using in vivo polyhydroxybutyrate (PHB) matrix association. *Protein Sci* 14:1387–1395
- Bastings MM, van Baal I, Meijer EW, Merkx M (2008) One-step refolding and purification of disulfide-containing proteins with a C-terminal MESNA thioester. *BMC Biotechnol* 8:76
- Belfort M, Derbyshire V, Stoddard BL, Wood DW (2005) Homing endonucleases and inteins. Springer, Berlin Heidelberg New York
- Billich A, Zocher R (1987) Enzymatic synthesis of cyclosporin A. *J Biol Chem* 262:17258–17259
- Caspi J, Amitai G, Belenkiy O, Pietrovski S (2003) Distribution of split DnaE inteins in cyanobacteria. *Mol Microbiol* 50:1569–1577
- Charalambous A, Andreou M, Skourides PA (2009) Intein-mediated site-specific conjugation of Quantum Dots to proteins in vivo. *J Nanobiotechnology* 7:9
- Cheriyian M, Perler FB (2009) Protein splicing: a versatile tool for drug discovery. *Adv Drug Deliv Rev* 61:899–907
- Chevalier BS, Stoddard BL (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucl Acids Res* 29:3757–3774
- Choi JI, Lee SY, Han K (1998) Cloning of the *Alcaligenes latus* polyhydroxyalkanoate biosynthesis genes and use of these genes for enhanced production of Poly(3-hydroxybutyrate) in *Escherichia coli*. *Appl Environ Microbiol* 64:4897–4903
- Choi JJ, Nam KH, Min B, Kim SJ, Söll D, Kwon ST (2006) Protein trans-splicing and characterization of a split family B-type DNA polymerase from the hyperthermophilic archaeal parasite *Nanoarchaeum equitans*. *J Mol Biol* 356:1093–1106
- Chong S, Xu MQ (1997) Protein splicing of the *Saccharomyces cerevisiae* VMA intein without the endonuclease motifs. *J Biol Chem* 272:15587–15589
- Chong S, Xu MQ (2005) Harnessing inteins for protein purification and characterization. In: Belfort M, Derbyshire V, Stoddard BL, Wood DW (eds) Homing endonucleases and inteins, vol 16. Springer, Berlin Heidelberg New York, pp 273–292
- Chong S, Mersha FB, Comb DG, Scott ME, Landry D, Vence LM, Perler FB, Benner J, Kucera RB, Hirvonen CA, Pelletier JJ, Paulus H, Xu MQ (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* 192:271–281
- Craik DJ (2006) Chemistry. Seamless proteins tie up their loose ends. *Science* 311:1563–1564
- Craik DJ, Daly NL, Bond T, Wayne C (1999) Plant cyclotides: a unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif. *J Mol Biol* 294:1327–1336
- Dassa B, Amitai G, Caspi J, Schueler-Furman O, Pietrovski S (2007) Trans protein splicing of cyanobacterial split inteins in endogenous and exogenous combinations. *Biochemistry* 46:322–330
- Dassa B, London N, Stoddard BL, Schueler-Furman O, Pietrovski S (2009) Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucl Acids Res* 37:2560–2573
- Dawson PE, Kent SB (2000) Synthesis of native proteins by chemical ligation. *Annu Rev Biochem* 69:923–960

- Derbyshire V, Wood DW, Wu W, Dansereau JT, Dalgaard JZ, Belfort M (1997) Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution. *Proc Natl Acad Sci U S A* 94:11466–11471
- Ding Y, Xu MQ, Ghosh I, Chen X, Ferrandon S, Lesage G, Rao Z (2003) Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *J Biol Chem* 278:39133–39142
- Douglas SE, Penny SL (1999) The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved syntenic groups confirm its common ancestry with red algae. *J Mol Evol* 48:236–344
- Driscoll DM, Copeland PR (2003) Mechanism and regulation of selenoprotein synthesis. *Ann Rev Nutr* 23:17–40
- Elleuche S, Pöggeler S (2007) Trans-splicing of an artificially split fungal mini-intein. *Biochem Biophys Res Commun* 355:830–834
- Elleuche S, Pöggeler S (2009) Inteins—selfish elements in fungal genomes. In: Anke T, Weber D (eds) *The mycota XV. Physiology and genetics. Selected basic and applied aspects*. Springer, Berlin Heidelberg New York, pp 41–61
- Elleuche S, Nolting N, Pöggeler S (2006) Protein splicing of PRP8 mini-inteins from species of the genus *Penicillium*. *Appl Microbiol Biotechnol* 72:959–967
- Evans TC Jr, Xu MQ (1999) Intein-mediated protein ligation: harnessing nature's escape artists. *Biopolymers* 51:333–342
- Evans TC Jr, Benner J, Xu MQ (1999) The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *J Biol Chem* 274:18359–18363
- Floss DM, Schallau K, Rose-John S, Conrad U, Scheller J (2009) Elastin-like polypeptides revolutionize recombinant protein expression and their biomedical application. *Trends Biotechnol* 28:37–45
- Fong BA, Wu WY, Wood DW (2009) Optimization of ELP-intein mediated protein purification by salt substitution. *Protein Expr Purif* 66:198–202
- Fong BA, Wu WY, Wood DW (2010) The potential role of self-cleaving purification tags in commercial-scale processes. *Trends Biotechnol* 28:272–279
- Ge X, Yang DS, Trabbic-Carlson K, Kim B, Chilkoti A, Filipe CD (2005) Self-cleavable stimulus responsive tags for protein purification without chromatography. *J Am Chem Soc* 127:11228–11229
- Georgiou G, Jeong KJ (2005) Proteins from PHB granules. *Protein Sci* 14:1385–1386
- Gillies AR, Hsui JF, Oak S, Wood DW (2008) Rapid cloning and purification of proteins: gateway vectors for protein purification by self-cleaving tags. *Biotechnol Bioeng* 101:229–240
- Gillies AR, Mahmoud RB, Wood DW (2009) PHB-intein-mediated protein purification strategy. *Methods Mol Biol* 498:173–183
- Gogarten JP, Senejani AG, Zhaxybayeva O, Olendzenski L, Hilario E (2002) Inteins: structure, function, and evolution. *Annu Rev Microbiol* 56:263–287
- Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ (1997) Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell* 91:85–97
- Hirata R, Ohsumi Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y (1990) Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J Biol Chem* 265:6726–6733
- Hondal RJ (2009) Using chemical approaches to study selenoproteins—focus on thioredoxin reductases. *Biochim Biophys Acta* 1790:1501–1512
- Johnson MA, Southworth MW, Herrmann T, Brace L, Perler FB, Wüthrich K (2007) NMR structure of a KlbA intein precursor from *Methanococcus jannaschii*. *Protein Sci* 16:1316–1328
- Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebel M, Stevens TH (1990) Protein splicing converts the yeast *TFPI* gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* 250:651–657
- Klabunde T, Sharma S, Telenti A, Jacobs WRJ, Sacchettini JC (1998) Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nat Struct Biol* 5:31–36
- Koonin EV (1995) A protein splice-junction motif in hedgehog family proteins. *Trends Biochem Sci* 20:141–142
- Liu XQ (2000) Protein-splicing intein: genetic mobility, origin, and evolution. *Ann Rev Genet* 34:61–76
- Liu XQ, Yang J (2003) Split *dnaE* genes encoding multiple novel inteins in *Trichodesmium erythraeum*. *J Biol Chem* 278:26315–26318
- Liu XQ, Yang J, Meng Q (2003) Four inteins and three group II introns encoded in a bacterial ribonucleotide reductase gene. *J Biol Chem* 278:46828–46831
- Liu JR, Duan CH, Zhao X, Tzen JT, Cheng KJ, Pai CK (2008) Cloning of a rumen fungal xylanase gene and purification of the recombinant enzyme via artificial oil bodies. *Appl Microbiol Biotechnol* 79:225–233
- Luo J, Hall BD (2007) A multistep process gave rise to RNA polymerase IV of land plants. *J Mol Evol* 64:101–112
- Mathys S, Evans TC, Chute IC, Wu H, Chong S, Benner J, Liu XQ, Xu MQ (1999) Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene* 231:1–13
- Meyer DE, Chilkoti A (1999) Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nat Biotechnol* 17:1112–1115
- Mills KV, Lew BM, Jiang S, Paulus H (1998) Protein splicing *in trans* by purified N- and C-terminal fragments of the *Mycobacterium tuberculosis* RecA intein. *Proc Natl Acad Sci USA* 95:3543–3548
- Mootz HD, Muir TW (2002) Protein splicing triggered by a small molecule. *J Am Chem Soc* 124:9044–9045
- Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A* 95:6705–6710
- Noren CJ, Wang J, Perler FB (2000) Dissecting the chemistry of protein splicing and its applications. *Angew Chem Int Ed Engl* 39:450–466
- Oeemig JS, Aranko AS, Djupsjobacka J, Heinamaki K, Iwai H (2009) Solution structure of DnaE intein from *Nostoc punctiforme*: structural basis for the design of a new split intein suitable for site-specific chemical modification. *FEBS Lett* 583:1451–1456
- Ozawa T, Umezawa Y (2005) Inteins for split-protein reconstitutions and their applications. In: Belfort M, Derbyshire V, Stoddard BL, Wood DW (eds) *Homing endonucleases and inteins*, vol. 16. Springer, Berlin Heidelberg New York, pp 307–323
- Ozawa T, Sako Y, Sato M, Kitamura T, Umezawa Y (2003) A genetic approach to identifying mitochondrial proteins. *Nat Biotechnol* 21:287–293
- Ozawa T, Nishitani K, Sako Y, Umezawa Y (2005) A high-throughput screening of genes that encode proteins transported into the endoplasmic reticulum in mammalian cells. *Nucleic Acids Res* 33:e34
- Paulus H (2000) Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* 69:447–496
- Perler FB (1998) Protein splicing of inteins and hedgehog autoproteolysis: structure, function, and evolution. *Cell* 92:1–4
- Perler FB (2002) InBase: the Intein Database. *Nucl Acids Res* 30:383–384
- Perler FB, Davis EO, Dean GE, Gimble FS, Jack WE, Neff N, Noren CJ, Thorner J, Belfort M (1994) Protein splicing elements: inteins and exteins—a definition of terms and recommended nomenclature. *Nucl Acids Res* 22:1125–1127

- Perler FB, Olsen GJ, Adam E (1997) Compilation and analysis of intein sequences. *Nucl Acids Res* 25:1087–1093
- Petrokovski S (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci* 3:2340–2350
- Petrokovski S (1998) Identification of a virus intein and a possible variation in the protein-splicing reaction. *Curr Biol* 8:R634–R635
- Poulter RT, Goodwin TJ, Butler MI (2007) The nuclear-encoded inteins of fungi. *Fungal Genet Biol* 44:153–179
- Saleh L, Perler FB (2006) Protein splicing in cis and in trans. *Chem Rec* 6:183–193
- Sancheti H, Camarero JA (2009) “Splicing up” drug discovery. Cell-based expression and screening of genetically-encoded libraries of backbone-cyclized polypeptides. *Adv Drug Deliv Rev* 61:908–917
- Scott CP, Abel-Santos E, Wall M, Wahn DC, Benkovic SJ (1999) Production of cyclic peptides and proteins in vivo. *Proc Natl Acad Sci U S A* 96:13638–13643
- Severinov K, Muir TW (1998) Expressed protein ligation, a novel method for studying protein-protein interactions in transcription. *J Biol Chem* 273:16205–16209
- Seyedsayamdost MR, Yee CS, Stubbe J (2007) Site-specific incorporation of fluorotyrosines into the R2 subunit of *E. coli* ribonucleotide reductase by expressed protein ligation. *Nat Protoc* 2:1225–1235
- Sharma SS, Chong S, Harcum SW (2006) Inteин-mediated protein purification of fusion proteins expressed under high-cell density conditions in *E. coli*. *J Biotechnol* 125:48–56
- Shen Y, Ai HX, Song R, Liang ZN, Li JF, Zhang SQ (2010) Expression and purification of moricin CM4 and human beta-defensins 4 in *Escherichia coli* using a new technology. *Microbiol Res* in press
- Shingledecker K, Jiang SQ, Paulus H (1998) Molecular dissection of the *Mycobacterium tuberculosis* RecA intein: design of a minimal intein and of a trans-splicing system involving two intein fragments. *Gene* 207:187–195
- Singleton SF, Simonette RA, Sharma NC, Roca AI (2002) Inteин-mediated affinity-fusion purification of the *Escherichia coli* RecA protein. *Protein Expr Purif* 26:476–488
- Southworth MW, Adam E, Panne D, Byer R, Kautz R, Perler FB (1998) Control of protein splicing by intein fragment reassembly. *EMBO J* 17:918–926
- Southworth MW, Benner J, Perler FB (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *EMBO J* 19:5019–5026
- Srinivasa Babu K, Muthukumaran T, Antony A, Prem Singh Samuel SD, Balamurali M, Murugan V, Meenakshisundaram S (2009) Single step intein-mediated purification of hGMCSF expressed in salt-inducible *E. coli*. *Biotechnol Lett* 31:659–664
- Starokadomskyy PL (2007) Protein splicing. *Mol Biol (Mosk)* 41:314–330
- Sun W, Yang J, Liu XQ (2004) Synthetic two-piece and three-piece split inteins for protein trans-splicing. *J Biol Chem* 279:35281–35286
- Sun P, Ye S, Ferrandon S, Evans TC, Xu MQ, Rao Z (2005) Crystal structures of an intein from the split *dnaE* gene of *Synechocystis* sp. PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing. *J Mol Biol* 353:1093–1105
- Tavassoli A, Benkovic SJ (2007) Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in *E. coli*. *Nat Protoc* 2:1126–1133
- Tavassoli A, Naumann TA, Benkovic SJ (2005) Production of cyclic proteins and peptides. In: Belfort M, Derbyshire V, Stoddard BL, Wood DW (eds) *Homing endonucleases and inteins*, vol. 16. Springer, Berlin Heidelberg New York, pp 293–305
- Terpe K (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 60:523–533
- Thali M (1995) Cyclosporins: immunosuppressive drugs with anti-HIV-1 activity. *Mol Med Today* 1:287–291
- Tori K, Dassa B, Johnson MA, Southworth MW, Brace LE, Ishino Y, Petrokovski S, Perler FB (2010) Splicing of the mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate block F nucleophile. *J Biol Chem* 285:2515–2526
- Turnmel M, Brouard J-S, Gagnon C, Otis C, Lemieux C (2008) Deep division in the Chlorophyceae (Chlorophyta) revealed by chloroplast phylogenomic analyses. *J Phycol* 44:739–750
- Van Roey P, Pereira B, Li Z, Hiraga K, Belfort M, Derbyshire V (2007) Crystallographic and mutational studies of *Mycobacterium tuberculosis* recA mini-inteins suggest a pivotal role for a highly conserved aspartate residue. *J Mol Biol* 367:162–173
- Wang S, Liu XQ (1997) Identification of an unusual intein in chloroplast ClpP protease of *Chlamydomonas eugametos*. *J Biol Chem* 272:11869–11873
- Waugh DS (2005) Making the most of affinity tags. *Trends Biotechnol* 23:316–320
- Weber G, Schörgendorfer K, Schneider-Scherzer E, Leitner E (1994) The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Curr Genet* 26:120–125
- Wieczorek R, Pries A, Steinbüchel A, Mayer F (1995) Analysis of a 24-kilodalton protein associated with the polyhydroxyalkanoic acid granules in *Alcaligenes eutrophus*. *J Bacteriol* 177:2425–2435
- Wood DW, Wu W, Belfort G, Derbyshire V, Belfort M (1999) A genetic system yields self-cleaving inteins for bioseparations. *Nat Biotechnol* 17:889–892
- Wood DW, Harcum SW, Belfort G (2005) Industrial applications of intein technology. In: Belfort M, Derbyshire V, Stoddard BL, Wood DW (eds) *Homing endonucleases and inteins*, vol. 16. Springer, Berlin Heidelberg New York, pp 345–364
- Wu H, Hu Z, Liu XQ (1998) Protein trans-splicing by a split intein encoded in a split *DnaE* gene of *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci U S A* 95:9226–9231
- Wu WY, Mee C, Califano F, Banki R, Wood DW (2006) Recombinant protein purification by self-cleaving aggregation tag. *Nat Protoc* 1:2257–2262
- Xu MQ, Evans TCJ (2001) Inteин-mediated ligation and cyclization of expressed proteins. *Methods* 24:257–277
- Zettler J, Schütz V, Mootz HD (2009) The naturally split *Npu* DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett* 583:909–914
- Zhao Z, Lu W, Dun B, Jin D, Ping S, Zhang W, Chen M, Xu MQ, Lin M (2008) Purification of green fluorescent protein using a two-intein system. *Appl Microbiol Biotechnol* 77:1175–1180
- Züger S, Iwai H (2005) Inteин-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. *Nat Biotechnol* 23:736–740